图书情报工作
第65卷 第10期 2021年5月

融合评论主题识别与技术属性多维度分析的技术机会发现研究*

■ 吴一平　白如江　刘明月　王效岳
山东理工大学信息管理研究院　淄博 255049

摘　要：[目的/意义]提出一种融合评论主题识别与技术属性多维度分析的技术机会发现方法，从技术需求驱动视角识别技术机会，为企业前瞻布局研发方向与进行科研管理规划提供决策建议支持。[方法/过程]以产品在线评论为研究数据源，首先，利用LDA主题模型识别出评论技术主题，提出技术评论主题强度和主题新颖度两个指标，筛选出新兴重点技术评论主题。然后，从学术论文、技术专利中人工选取技术属性词，通过TF-IDF值计算得到评论高频词，结合专家知识进一步筛选出技术特征词，构建产品技术属性词-技术特征词表。通过相关性计算分别得到与评论相关和与新兴重点技术评论主题相关的技术属性。最后，提出一种产品重要技术属性识别指标模型并设计一种多维度分析方法，分析产品重要技术属性的特征情况，最终识别出蕴含在评论文本中的新兴技术机会。[结果/结论]实验结果表明该方法能够有效地识别技术机会，为企业产品技术研发管理提供参考。

关键词：技术机会发现　技术属性分析　主题识别　评论挖掘
分类号：G251
DOI：10.13266/j.issn.0252-3116.2021.10.007

1　引言

目前，我国企业正面临着全球技术革命与科技竞争浪潮等严峻形势。推进前瞻性的技术机会发现研究工作，有助于企业夺取未来市场竞争中的先发优势，并提供用以支撑重要科技创新决策与制定核心研发战略的重要情报依据。为此，研究人员通常使用学术论文和技术专利，从技术驱动的角度识别和监测新兴技术的发展趋势，却很少利用与新兴技术相关的社交媒体数据进行技术机会发现研究[1]。随着电子商务的蓬勃发展与网络购物的推广普及，电商平台上涌现出的大量方便易得、内容丰富的在线产品评论数据，对于获取技术需求反馈信息具有重要的研究价值。

那么，如何有效地利用在线评论，从中发现技术机会？如何基于技术机会的特征，对前瞻性技术机会进行分析、识别与筛选？为此，本研究提出一种融合评论主题识别与技术属性多维度分析的前瞻技术机会发现研究方法，将技术机会发现研究的重心前移到技术需求端，通过有效地萃取分析评论数据中的技术需求信息，以直接、准确地发现前瞻性技术机会。

2　相关研究进展

技术机会是推进技术创新活动的关键影响因素与重要决策参考，识别技术机会的能力是国家和企业的最重要的研发核心竞争力之一[2]。1995年，美国佐治亚理工学院教授A. Porter在《技术机会分析》(*Technology Opportunities Analysis*)一文中正式提出“技术机会”(technology opportunities)的概念，指通过对某领域内已有技术在竞争主体间横向对比和随时间纵向发展趋势及相互关系的挖掘，推断该领域即将可能出现的技术形态或技术发展点[3]，开辟了技术机会发现相关研究领域的先河。李保明从技术本身及经济学两个角

* 本文系山东省高等学校青创科技支持计划“科技大数据驱动的智慧决策支持创新团队-面向新旧动能转换的新兴科学研究前沿识别研究”(项目编号:2019RWG033)研究成果之一。
作者简介：吴一平(ORCID:0000-0001-9426-7328)，硕士研究生；白如江(ORCID:0000-0003-3822-8484)，研究馆员，硕士生导师，通讯作者，E-mail:brj@sdut.edu.cn；刘明月(ORCID:0000-0002-4335-9369)，硕士研究生；王效岳(ORCID:0000-0002-7100-7758)教授，博士，硕士生导师。
收稿日期:2020-11-30　修回日期:2021-01-23　本文起止页码:56-67　本文责任编辑:徐健

度对技术机会进行阐释，认为技术机会是技术进步的机会，是企业(或社会)提供的新技术成功应用于生产的可能性[4]。随着学界对技术机会发现研究的不断深入，其研究概念、研究数据源、研究方法也在不断丰富发展。

2.1 技术机会的概念内涵在情报学与企业管理等研究领域不断丰富

李保明[4]将技术机会划分为内涵的技术机会和外延的技术机会。其中，内涵的技术机会是指现存技术的规范或性能有改进的可能性；外延的技术机会是指一个特定的技术有转移到其他许多技术系统的可能性，且经转移后在很多功能上可以比正在应用中的技术系统更有效。陈震红等[5]认为技术机会是技术变化带来的创业机会。康宇航[6]认为技术机会是通过对某技术领域内已有技术发展趋势及相互关系的挖掘，发现最新技术动向，推断该领域可能出现的技术形态或技术发展点。G. Cecere 等[7]认为技术机会是企业持续创新过程中的核心要素。技术机会发现(Technology Opportunity Discovery，TOD)可以理解为一种在技术机会相关研究理论、方法和技术的指导下，主动挖掘含有技术机会信息的数据源，发现潜在的技术创新发展契机的研究活动。技术机会发现研究有助于把握最新技术动向，为国家宏观科技决策制定与企业技术研发创新管理提供情报参考。

2.2 技术机会发现研究所采用的数据源主要是论文与专利

A. Porter 提出技术机会发现与当前存在的技术密切相关并具有复杂的互动机制[3]。F. Malerba 等[8]提出，专利数据、科技期刊、科技报告等反映最新的技术信息和动态的科技资源，为挖掘潜在的技术创新机会提供了可能，为技术机会发现研究提供了研究数据源参考。李欣、黄鲁成等[9]指出，研究人员通常使用学术论文和专利数据从技术角度识别和监视新兴技术的趋势。科学论文是基础科学重要载体，技术专利是技术信息的重要载体。因此，目前技术机会发现研究数据源仍以科学论文与技术专利等数据源为主。

2.3 技术机会发现的研究方法随着时代发展不断丰富创新

传统的技术机会发现研究主要依靠基于专家知识，该方法在局部范围内或者细分技术领域能够保证较高的效率与准确性。大数据时代机器学习技术的蓬勃发展，为技术机会发现研究提供了充足的技术方法支撑，并且随着市场环境的激烈变化与技术创新周期的日趋缩短，技术机会发现研究方法也在实证研究中不断深入与丰富，包括但不限于综合应用基于专家知识的技术机会发现方法[10]、基于文献计量的技术机会发现方法[8,11－15]、基于文本挖掘的技术机会发现方法[16－18]、基于社会网络分析的技术机会发现方法等[8,12－15]。例如，M. Y. Wang 等[16]提出，科学与技术知识之间存在的差异具有挖掘潜在技术机会的可能，因此通过专利文本挖掘方法结合聚类算法，研究分析科学与技术知识之间的差距以发现潜在技术机会，并在微藻生物燃料领域进行实证研究。李欣、黄鲁成等[11]运用文献计量方法统计分析了染料敏化太阳能光伏技术产业的技术热点、技术前沿、技术机会、技术发展趋势等，构建了基于文献计量、专利分析和技术路线图研究方法的新兴技术产业未来发展分析框架，以客观数据统计结果发现新兴技术机会。王京安[15]通过对比分析物联网技术研究领域的科技论文与专利文献生成的关键词聚类网络图、作者聚类网络图、机构聚类网络图以及关键词聚类时间线网络图等，判断技术研究热点发现技术机会，揭示了物联网行业领域未来发展趋势。

虽然目前的技术机会发现研究已经取得了丰富的研究成果，但仍存在一些问题。在研究数据源上，一方面，目前主要选用论文、专利等记录已有技术的信息载体进行技术机会发现研究。所得研究结果往往具有一定时滞性，可能落后于技术领域的最新趋势，难以满足获得最具前瞻性的技术机会发现结果的研究需求。产品评论数据方便易得、实时更新，并且直接客观地反映了用户对产品技术的需求与感知，是对研究技术机会发现具有重要价值的科技数据。挖掘用户产品评论中的技术需求反馈能够更加直接、更为前瞻地发现未来技术机会。然而，目前的研究没有对产品在线评论引起足够重视，进而导致目前技术机会发现研究结果与用户直接技术需求结合方面存在一定局限性，从用户技术需求驱动视角的技术机会发现研究机理机制有待进一步探索。

在研究方法上，基于专家知识、文献计量、文本挖掘、社会网络分析等技术机会发现方法体系已经较为丰富完善，但是现有的实现技术与研究方法在基于产品在线评论数据的技术机会发现研究的识别效率、挖掘结果方面的精确性以及算法的适用性等方面，仍待进行充分的实证研究。目前的技术机会发现研究过程仍未充分结合时序、品牌型号等因素深入分析，因而研究结果往往没能全面地反映技术机会的时间敏感性与

品牌型号间的差异性等特征。此外，目前基于产品评论驱动的技术机会发现的相关研究的理论方法模型、技术实现路径以及实证研究成果整体比较有限，技术机会发现识别方法比较粗略笼统，有待基于更为科学严谨的研究方法与技术手段，将量化分析与内容分析相结合，从用户技术需求驱动视角深入技术机会发现研究。

因此，本研究提出一种融合评论主题识别与技术属性多维度分析的前瞻技术机会发现研究方法，以产品在线评论数据为研究数据源，通过新兴重点技术评论主题识别分析、产品重要技术属性识别分析以及技术属性多维度分析，发现产品评论中的前瞻性技术机会，为技术机会发现与科研创新管理提供理论参考与模型支撑。

3 研究思路

本研究基于电商平台智能手机产品在线评论、学术论文、科技报告、专家知识等数据源，通过研究分析评论数据中的用户技术需求反馈，驱动前瞻性技术机会发现。基于这样的研究前提，本研究的研究思路如图 1 所示。主要包括 5 部分内容，第一部分利用主题模型识别出评论文本中的新兴重点技术主题，第二部分构建了技术属性词和技术特征词表，第三部分在前面两部分研究基础上识别出与新兴重点技术主题相关的技术属性，第四部分识别出重点技术主题下具体产品的相关技术属性，第五部分提出一种产品重要技术属性多维度分析模型，最终识别出蕴含在评论文本中的技术机会。

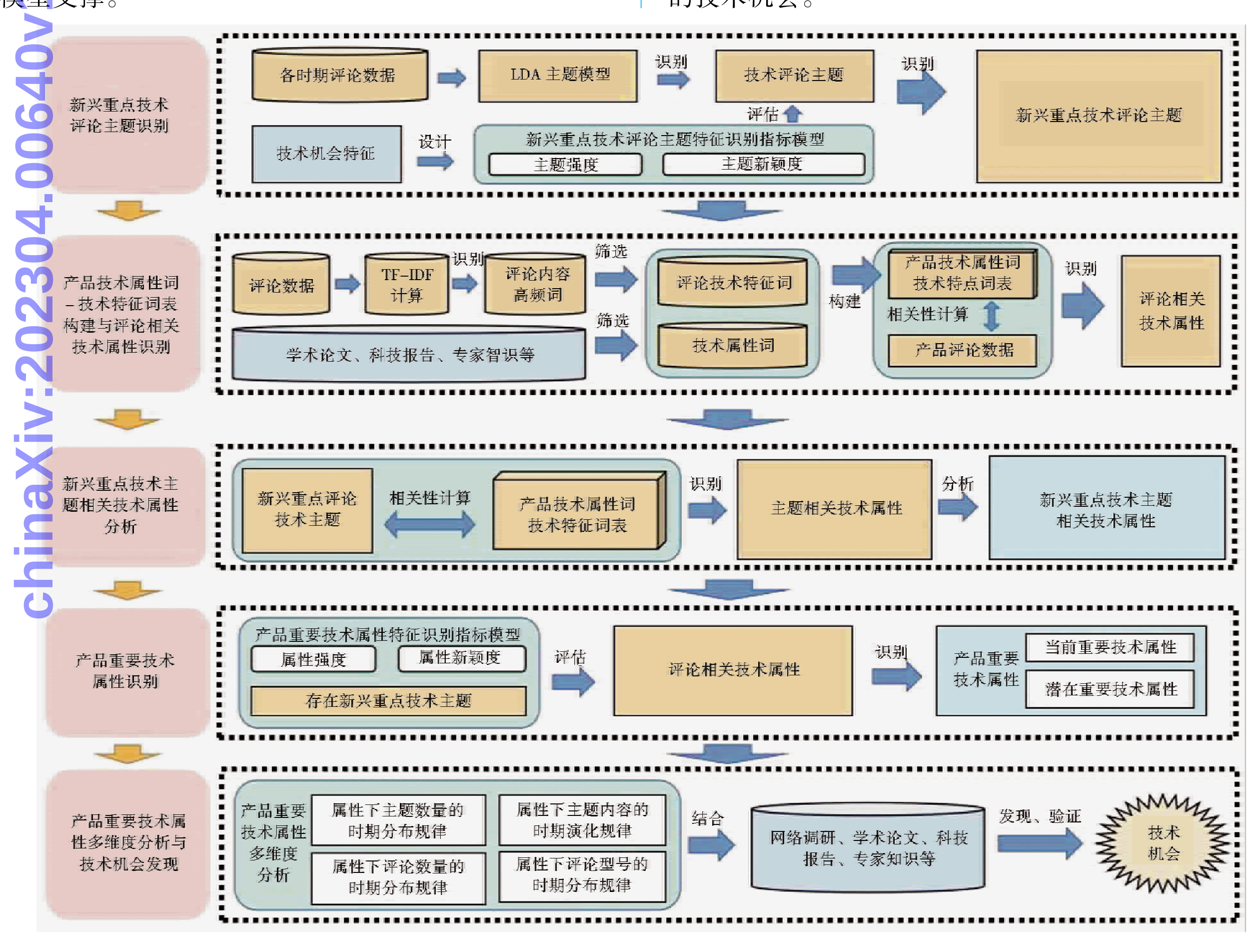


图 1 研究思路示意

3.1 新兴重点技术评论主题识别

3.1.1 基于 PLDA 主题模型的技术评论主题识别

本文基于并行潜在狄利克雷分布模型（Parallel Latent Dirichlet Allocation，PLDA）设计了产品在线评论主题识别方法，对评论数据进行时间分割与主题识别，旨在通过识别出各时期的评论主题及主题词，研究各时期评论数据中的技术内容。PLDA 主题模型利用吉布斯采样（Gibbs sampling）进行参数求解，提高了算法的运行效率和并行的加速比，能够高效准确地识别评论文本中的主题及其主题词[19]。

3.1.2 新兴重点技术评论主题识别

综合考虑要寻找的技术机会具有一定时效性与用

户需求性特征，本文设计了包含技术评论主题强度、主题新颖度两项指标的新兴重点技术评论主题识别模型。通过该指标模型计算，精确地识别出主题强度较大、新颖度较高的新兴重点技术评论主题。

（1）技术评论主题强度指标。该指标能够直观反映主题的受关注度与参与度。定义该指标为每个主题内部的评论数量占该时期评论总量的权重，计算公式为：

$$TI = \frac{X_i}{\sum_{j=1}^{n} X_j} \qquad \text{公式(1)}$$

其中，TI 代表某时期的技术评论主题强度（topic intensity）；

X_i 代表某时期识别出的某技术评论主题下的评论数量；

$\sum_{j=1}^{n} X_j$ 代表该时期所有技术评论主题下评论数量之和。

为了确定评论所属主题，确定主题下评论的数量，本研究利用计算余弦相似度的方法，得到评论与主题之间的相关性。

第一步，构建向量空间模型（Vector Space Model，VSM），把主题与评论用向量的方式进行描述，向量空间模型中用 Comment 表示评论、Topic 表示主题、L 表示评论信息词或主题词、w 表示主题词或属性特征词权重，主题向量可用主题词表示为 $Topic_i = \{L_1, L_2, L_3, \cdots\cdots L_m\}$、评论向量可用评论信息词表示为 $Comment_j = \{L_1, L_2, L_3, \cdots\cdots L_n\}$、主题词的权重向量为 Topic Vector = $\{W_1, W_2, W_3, \cdots\cdots W_n\}$，评论信息词的权重向量为 Comment Vector = $\{W_1, W_2, Wj_3, \cdots\cdots_n\}$，每个主题词或技术特征词都有一个权重。

第二步，计算评论与主题之间的相似度，计算结果介于[0,1]之间。参考余弦相似度计算方法[20]，设计了评论与主题相似度计算公式为：

$$Sim(Topic_i, Comment_j) = \cos\theta = \frac{\sum_{k=1}^{n} w_k(Topic_i) \times w_k(Comment_j)}{\sqrt{(\sum_{k=1}^{n} w_k^2(Topic_i)) \times (\sum_{k=1}^{n} w_k^2(Comment_j))}} \qquad \text{公式(2)}$$

其中，分子 $\sum_{k=1}^{n} w_k(Topic_i) \times w_k(Comment_j)$ 表示评论主题向量与评论信息词向量的点乘积；

分母 $\sqrt{(\sum_{k=1}^{n} w_k^2(Topic_i)) \times (\sum_{k=1}^{n} w_k^2(Comment_j))}$ 表示评论主题向量与评论信息词向量模的乘积；

$Sim(Topic_i, Comment_j)$ 是评论主题向量与评论信息词向量之间的相似度；

设定集合 A 为某主题下的评论集合，TC_Sim 为该主题下的某评论相似度，阈值为 α，若相似度大于 α 则评论属于该主题，否则不属于该主题。表示为 $A = \{TC_Sim \mid TC_Sim > \alpha, \alpha \in [0,1]\}$。

设置技术评论主题强度 S 的阈值为 β，若某主题强度大于 β，则说明该主题强度较高。

主题强度越高，表明用户对该主题的关注度与参与度高，则该主题可能是新兴重点技术评论主题。

（2）技术评论主题新颖度指标。该指标能够揭示技术评论主题随时间变化的发展趋势是新生、发展还是消亡。定义该指标计算公式为：

$$TN = \sum_{i=1}^{n} \frac{date_i}{total_num} \qquad \text{公式(3)}$$

其中，TN 代表技术评论主题新颖度（topic novelty）；

total_num 为该技术评论主题下的评论数量之和；

$Date_i$ 为该技术评论主题下每篇评论发表日期；

$\sum_{i=1}^{n} \frac{date_i}{total_num}$ 计算结果为该主题下全部评论的平均发表日期，即技术评论主题新颖度。计算得到的数值越大则主题新颖度越强，表明该评论主题的内容发表的时间越新。

设定主题新颖度 N 的阈值为 γ，若主题新颖度大于 γ，则表明布局年份新，该主题可能是新兴重点技术评论主题。

设定集合 B 为新兴重点技术评论主题集合，设定技术主题强度 S 高于阈值 β，主题新颖度 N 高于阈值 γ 的主题为新兴重点技术评论主题，属于集合 B，表示为 $B = \{S, N \mid S > \beta, N > \gamma, \beta \in R, \gamma \in R\}$。

3.2 产品技术属性词－技术特征词表构建与评论相关技术属性识别

3.2.1 产品技术属性词－技术特征词表构建

为了后续识别出新兴重点主题包含哪些技术属性，本文提出产品技术属性词－技术特征词表构建方法。

第一步：对产品在线评论数据进行文本预处理与 TF-IDF 计算，识别评论内容高频词，并结合专家知识，进一步筛选出技术特征词。

第二步：通过整合学术论文、专利信息与专家报告，筛选较为权威的技术属性词。

第三步：根据第二步筛选出的权威技术属性词，与第一步筛选出的技术特征词进行属性－技术词匹配。例如：对技术属性词“声音”进行技术特征词匹配，优先从 TF-IDF 高频词中筛选含有“声”“响”

“音”“噪”“铃”等能够表征“声音”的词语，并从中进一步筛选合适的技术特征词，将其分类到“声音”技术属性下。构建完成后通过专家进一步分析判读，最终形成产品技术属性词－技术特征词表，如表 1 所示：

表 1 产品技术属性词－技术特征词表（部分）

技术属性序号	技术属性词	技术特征词
TECH_1	声音	音质、响、音量、声音洪亮、声音、通话音质、音响、外放、听歌、震耳欲聋、听筒、音色、柔音、无噪音……
TECH_2	游戏	手机游戏、打游戏、小游戏、玩游戏、游戏、可玩性、王者、团战、泡泡龙、网络游戏、火线……
……	……	……

3.2.2 评论相关技术属性识别

为了确定评论所属技术属性，确定技术属性下评论的数量，本研究利用计算评论与产品技术属性词－技术特征词表之间余弦相似度的方法，参见公式(2)，得到产品评论与评论属性之间的相关性。设定集合 C 为某属性下的评论集合，AC_Sim 为属性与评论的相似度，阈值为 δ，若相似度大于 δ 则评论属于该技术属性，否则不属于某项技术属性。表示为 $C = \{AC_Sim \mid AC_Sim > \delta, \delta \in [0,1]\}$。

3.3 新兴重点技术评论主题相关技术属性分析

为了确定新兴重点技术评论主题所属技术属性，明确技术机会发现范围。本研究设计了基于余弦相似度的主题与技术属性相关性计算方法，识别评论主题的相关技术属性，参见公式(2)。设定集合 D 为某属性下的主题集合，AT_Sim 为属性与主题的相似度，阈值为 ε，若相似度大于 ε 则主题属于该技术属性，否则不属于某项技术属性。表示为 $D = \{AT_Sim \mid AT_Sim > \varepsilon, \varepsilon \in [0,1]\}$。

3.4 产品重要技术属性识别

为了提高产品重要技术属性识别结果的精确度，本文设计了一种基于余弦相似度的评论与属性相似度计算方法，并构建了一种包含技术属性强度、技术属性新颖度两项指标的产品重要技术属性特征识别指标模型。使用构建的产品重要技术属性特征识别指标对技术属性进行评估，并通过确认该技术属性中包含新兴重点技术评论主题，识别产品重要技术属性（包含当前重要技术属性与潜在重要技术属性）。对识别出的产品重要技术属性进行技术属性内涵分析、属性强度分析、属性新颖度分析，从而研究产品重要技术属性的技术领域内容、属性受关注程度以及属性时期发展情况。

3.4.1 评论技术属性强度指标

该指标直观反映技术属性的受用户关注度与评论参与度情况。定义每个属性内部的评论数量占该时期评论总量的权重，评论属性强度计算公式为：

$$AI = \frac{C_i}{\sum_{j=1}^{n} C_j} \qquad \text{公式(4)}$$

其中，AI 代表某时期的评论技术属性强度（attribute intensity）；

C_i 代表某时期识别出的某评论技术属性下的评论数量；

$\sum_{j=1}^{n} C_j$ 代表该时期所有评论技术属性下评论数量之和。

属性强度越高，表明用户对该属性的关注度与评论参与度高，该属性可能是产品重要技术属性。

设定评论技术属性强度的阈值为 Q，若某属性强度大于 Q，则说明该属性强度较高，可能是包含技术机会产品的重要技术属性。

3.4.2 评论技术属性新颖度指标

该指标能够反映技术属性的评论年份布局情况，揭示该技术属性受关注的时期特征。设计评论技术属性新颖度计算公式为：

$$AN = \sum_{i=1}^{n} \frac{date_i}{total_num} \qquad \text{公式(5)}$$

其中，AN 代表评论技术属性新颖度（attribute novelty）；

total_num 为该评论技术属性下的评论数量之和；

$date_i$ 为该技术评论评论属性下每篇评论发表日期；

$\sum_{i=1}^{n} \frac{date_i}{total_num}$ 计算结果为该属性全部评论的平均发表日期，即评论技术属性新颖度。

属性新颖度越强，则表明该评论属性的内容发表的时间越新。

设定属性新颖度的阈值为 X，若属性新颖度大于 X，则表明该属性布局年份新，可能是包含技术机会的产品重要技术属性。

设定集合 E 为产品当前重要技术属性集合，设定评论技术属性强度 Q 高于阈值 ζ，属性新颖度 X 高于阈值 η 的且包含新兴重点技术评论主题的技术属性为产品当前重要技术属性，属于集合 E，表示为 $E = \{Q, X, AT_Sim \mid Q > \zeta, X > \eta, AT_Sim > \varepsilon, \zeta \in R, \eta \in R, \varepsilon \in [0,1]\}$。

设定集合 F 为潜在重要技术评论主题集合，设定评

论技术属性强度 Q 低于阈值 ζ，属性新颖度 X 高于阈值 η 的且包含新兴重点技术评论主题的技术属性为潜在重要技术属性，属于集合 E，表示为 $F=\{Q,X,AT_Sim \mid Q<\zeta, X>\eta, AT_Sim>\varepsilon, \zeta\in R, \eta\in R, \varepsilon\in[0,1]\}$。

设定集合 G 为产品重要技术属性集合，$G=E\cup F$，$E\subseteq G, F\subseteq G$。

3.5 产品重要技术属性多维度分析与技术机会发现

为了提高技术机会发现结果的全面性与精确性，本研究设计了产品重要技术属性多维度分析与技术机会发现方法，对识别出的产品重要技术属性进行技术属性下的主题数量时期分布规律分析、技术属性下的主题内容时期演化规律分析、技术属性下的评论数量时期分布规律分析以及技术属性下的评论型号时期分布规律分析，最终结合网络调研、学术论文、科技报告、专家知识等综合调研分析结果，发现技术机会。

技术属性下的主题数量时期分布规律分析，旨在揭示用户在不同时期对于该技术属性的反馈与关注广度等情况。

技术属性下的主题内容时期演化规律分析，旨在揭示用户对该技术属性反馈内容的时间变化。

为了计算属性下评论主题之间的相似度，参见公式(2)，计算结果介于[0,1]之间。

设定主题之间相似度为 TT_Sim，阈值为 θ，若相似度大于 θ 则两个主题互为相似评论主题，否则互为不同评论主题。集合 F 为相似评论主题集合，表示为 $F=\{TT_Sim \mid TT_Sim>\theta, \theta\in[0,1]\}$。

技术属性下的评论数量时期分布规律分析，旨在揭示用户对该技术属性的评论参与度与用户关注度的时期分布变化规律。

技术属性下的评论型号时期分布规律分析，旨在揭示不同时期用户对不同型号品牌的产品的技术反馈和诉求分布规律。

4 基于产品评论数据驱动的技术机会发现实证研究

本文以京东电商平台产品评论数量排名前 14 位的智能手机评论数据为数据源，数据源采用京东平台提供的 json 格式数据，以 Apple iPhone 11 为例，其 URL 为：https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment98&productId=100008348542&score=0&sortType=5&page=0&pageSize=50&isShadowSku=0&fold=1。使用 Python 构建爬虫程序采集数据，爬虫时间：2020 年 4 月 14 日，共得到 13 870 条评论数据，通过评论有用性筛选与去除重复项后共有 12 889 条评论数据。具体抓取产品、评论数量和评论时间分布情况如表 2 所示：

表 2 京东电商平台产品评论数量排名前 14 位的智能手机评论数据

序号	品牌	爬取评论数量(条)
1	Apple iPhone 11	990(2019:378,2020:612)
2	Apple iPhone 8 Plus	22(2017:3,2019:5,2020:14)
3	Apple iPhone XR	990(2018:116,2019:593,2020:281)
4	Redmi 8A	990(2019:192,2020:798)
5	redmi note 8 pro	990(2018:370,2019:620)
6	vivo Z5x	990(2019:656,2020:334)
7	华为 nova5 Pro	990(2019:578,2020:412)
8	华为 P30	990(2019:545,2020:445)
9	华为 P30Pro	990(2019:647,2020:343)
10	荣耀 20s	990(2019:251,2020:739)
11	荣耀 9x	990(2019:477,2020:513)
12	荣耀 v20	990(2018:2,2019:776,2020:212)
13	荣耀畅玩 7	987(2018:464,2019:295,2020:228)
14	小米 note8	990(2019:359,2020:631)

该实验环境是选用 Python、百度 Aistudio、数据挖掘软件 KNIME、Excel 等平台和软件进行数据处理与分析。

4.1 PLDA 主题识别与新兴重点技术评论主题识别研究

对评论数据进行时间分割，结合各时期具体评论数量，共将数据集划分 5 个时间序列组，分别对每组评论数据进行 PLDA 主题识别，共识别出 50 个评论主题，每个主题下 15 个关键词，主题关键词由 PLDA 主题模型根据关键词在主题中出现的概率大小自动生成。

为方便数据统计和分析，并结合各时期具体评论收集情况，本研究对主题进行统一命名，2017-2018 年为时期Ⅰ，2019 年 1 月-2019 年 4 月为时期Ⅱ，2019 年 5 月-2019 年 8 月为时期Ⅲ，2019 年 9 月-2019 年 12 月为时期Ⅳ，2020 年 1 月至 4 月为时期Ⅴ。将部分主题及主题词识别结果以矩阵形式表示，部分结果见表 3。

本文通过所设计的包含主题强度、主题新颖度两项指标的新兴重点技术评论主题特征识别指标模型计算，识别新兴重点技术评论主题。设定主题强度阈值

表 3 主题 – 主题词矩阵(部分)

主题编号	主题内容
Ⅰ_topic_0	屏幕 信号 效果 没 很大 8p 不 电池 网上 使用 一天 边框
Ⅰ_topic_1	屏幕 屏 采用 畅玩 7 拍 智能 全面 体验 电池 像素 震撼
Ⅰ_topic_2	智能 屏 全面 7 功能 畅玩 买 通话 环境 摄像头 荣耀 听筒
Ⅰ_topic_3	显示 现在 比较 买 去 不 效果 使用 手 小时 加上 来说
Ⅰ_topic_4	买 屏幕 速度 拍照 不 效果 苹果 很多 11 双 很快 运行
Ⅰ_topic_5	买 好评 华为 速度 老人 性价比 很快 手感 快递 价格 运行
Ⅰ_topic_6	买 不 X 解锁 xr 扬声器 面容 屏 屏幕 全面 苹果 很大 续航
Ⅰ_topic_7	不 买 不用 微信 数据 软件 屏幕 xs 点 续航 价位 流量
Ⅰ_topic_8	速度 很快 屏幕 拍照 包装 使用 整体 物流 一次 购物 很多
Ⅰ_topic_9	外观 买 老人 物流 屏 收到 全面 手感 屏幕 购买 性价比
Ⅱ_topic_0	买 屏幕 屏 很快 问题 速度 拍照 没 苹果 不是 之前 xr
Ⅱ_topic_1	3 不 屏幕 使用 电量 拓展 坚果 运行 买 功能 快充 V20
Ⅱ_topic_2	华为 买 屏幕 苹果 体验 拍照 不 使用 外观 没 荣耀 细节
Ⅱ_topic_3	屏幕 速度 拍照 华为 运行 效果 识别 电池 续航 不 模式
Ⅱ_topic_4	机器 拍照 屏幕 效果 到手 没 吃 不 鸡 买 电池 华为 抢购
Ⅱ_topic_5	买 苹果 使用 再 不 很快 物流 值得 速度 屏幕 问题 性价比
……	……

为 0.08,主题新颖度阈值为 2019,设定主题强度、主题新颖度高于阈值的主题即为新兴重点技术评论主题,共识别出 19 个新兴重点技术评论主题,部分结果见表 4。新兴重点技术评论主题中,除了包含对技术属性与产品特征的描述,还体现了对老人、长辈等用户群体的使用感以及对产品性价比的关注。

表 4 新兴重点技术评论主题(部分)

主题	主题强度	主题新颖度	主题词(部分)
Ⅲ_topic_3	0.31	2019.79	拍照、速度、屏幕、运行、外观、待机时间、音效、外形、特色、电池、手感……
Ⅳ_topic_5	0.18	2019.46	效果、拍照、速度、运行、屏幕、外观、待机时间、音效、外形、特色、手感……
Ⅴ_topic_8	0.15	2020.26	老人、性价比、妈妈、使用、内存、屏幕、适合、红米、价格、华为……

4.2 产品技术属性词 – 技术特征词表构建研究

通过 TF-IDF 计算从全部评论中提取出高频词,结合专家判读,从高频词中筛选出能够表征产品技术特征的词语,形成产品技术特征词表,基于专家知识,从论文、专利、报告等数据来源中抽象出技术属性词,共筛选出 13 种技术属性,并通过专家知识,匹配技术属性词与技术特征词,构建产品技术属性词 – 技术特征词表,部分结果见表 5。

4.3 新兴重点技术评论主题相关技术属性分析研究

通过计算新兴重点技术评论主题与产品技术属性词 – 技术特征词表之间的相似度,分析新兴重点技术评论主题相关技术属性。设定评论主题与技术属性的

表 5 产品技术属性词 – 技术特征词表(部分)

技术属性序号	技术属性词	技术特征词
TECH_6	摄像功能	拍照、自拍镜、拍照片、图像、晒图、拍不出、调焦照相、摄像机、变焦镜头、摄影、摄像、四摄、夜拍、变焦、反光、光圈……
TECH_7	存储	存储空间、存储量、显热、储量、内存、存储、内存卡、内存容量、运存够、内存不足……
TECH_8	电池	费电、电池容量、电池电量、电容量、蓄电量、耗电、续航力、余电、快充、续航、耗电量、电力、蓄电……

相似度阈值为 1%,采用 Echarts 可视化平台,将部分新兴重点技术评论主题与技术属性的相关情况以热力图的形式呈现。如图 2 所示(为保证图示效果,按相似度乘以 100 进行绘制),横轴为识别出的新兴重点技术评论主题,纵轴是本文构建的技术属性,着色方块代表主体与属性之间的相关性,颜色越深代表主题与属性相关性越强。同一主题可能与多种技术属性相关,其中近期与新兴重点技术评论主题相似度较高的技术属性主要有 TECH_8(电池)、TECH_3(触控/NFC/智能遥控技术)、TECH_7(存储)、TECH_10(附件)等,反映近期用户对这些技术属性关注度较高,在新兴重点技术评论主题中分布较广的技术属性主要有 TECH_1(声音)、TECH_8(电池)、TECH_12(外观设计)等,反映近期用户对这些技术属性关注度较广,可能存在技术机会。

4.4 基于多维度指标模型的产品重要技术属性识别分析研究

本研究使用构建的产品重要技术属性特征识别指标对技术属性进行评估,并通过确认该技术属性中包含新兴重点技术评论主题,识别当前重要技术属性与潜在重要技术属性两种产品重要技术属性。

参见公式(2)计算评论与产品技术属性词 – 技术特征词表之间的相似度,识别技术属性相关评论,设定评论与技术属性相似度阈值为 3%。通过使用公式(4)计算技术属性强度,使用公式(5)计算属性新颖度,设定属性强度阈值为 0.08,属性新颖度阈值为 2019.7。设定属性强度、属性新颖度高于阈值,并且与新兴重点技术评论主题相关的技术属性是产品重要技术属性中的当前重要技术属性,应该引起首要重视。设定属性强度低于阈值,属性新颖度高于阈值,且包含新兴重点技术评论主题的技术属性即为产品重要技术属性中的潜在重要技术属性。产品重要技术属性识别结果见表 6、表 7。

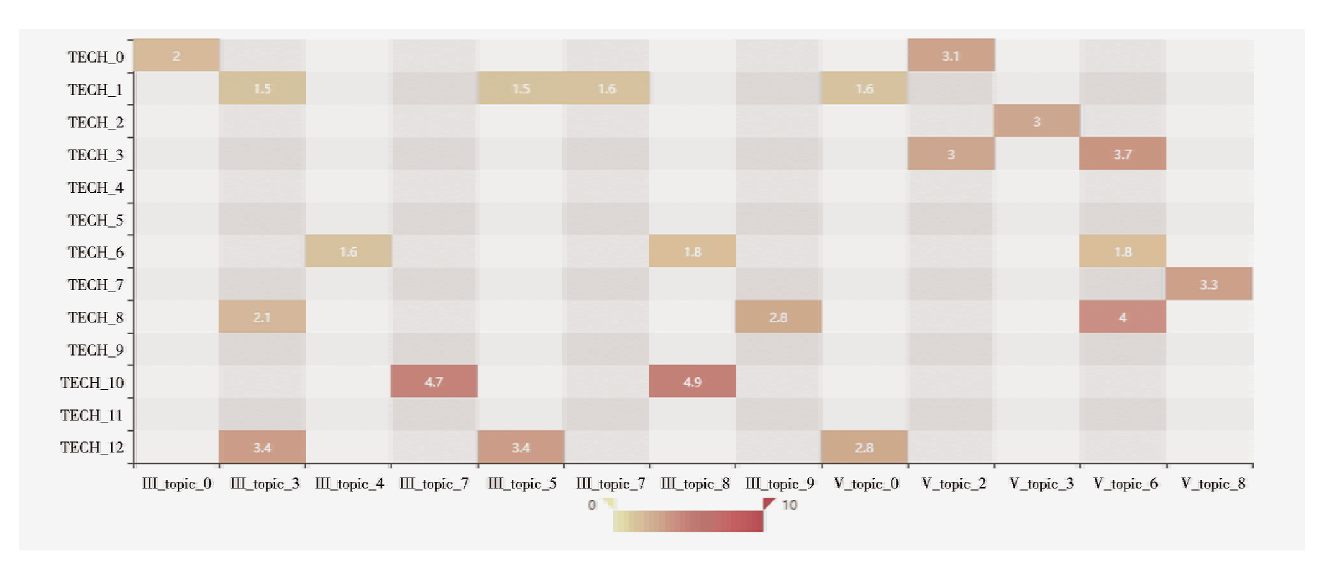


图2　新兴重点技术评论主题相关技术属性热力图

表6　产品当前重要技术属性

技术属性序号	技术属性	技术属性强度	技术属性新颖度
TECH_1	声音	0.082	2019.707
TECH_6	摄像功能	0.082	2019.711
TECH_7	存储	0.081	2019.72
TECH_8	电池	0.082	2019.704
TECH_10	附件	0.082	2019.712
TECH_12	外观设计	0.082	2019.720

表7　产品潜在重要技术属性

技术属性序号	技术属性	技术属性强度	技术属性新颖度
TECH_0	处理器/网络/数据传输技术	0.076	2019.759
TECH_2	游戏	0.075	2019.74
TECH_3	触控/NFC/智能遥控技术	0.078	2019.763

4.5　产品重要技术属性多维度分析方法与技术机会发现研究

4.5.1　产品重要技术属性下主题数量时期分布规律分析研究

参见公式(2)计算主题与产品技术属性词－技术特征词表之间的相似度。由于主题词的提取结果较为凝练，制定的技术属性－技术特征词表较为精确，设定主题与属性相似度以1%作为阈值，相似度高于阈值则认为是属性相关主题，产品重要技术属性下主题数量的时期分布情况如图3所示：

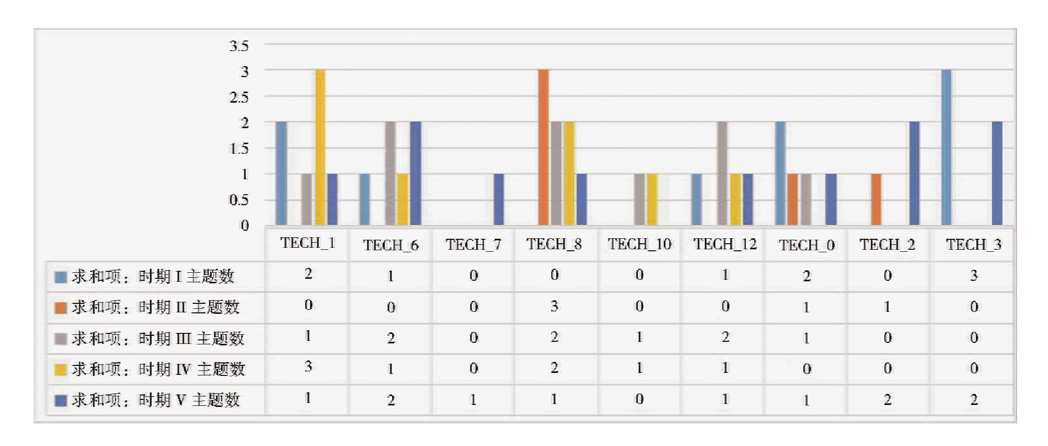


图3　产品重要技术属性下主题数量时期分布情况

从图3中可以看出，在Ⅰ时期，评论主题主要涉及TECH_3(触控/NFC/智能遥控技术)，其次是TECH_0(处理器/网络/数据传输技术)、TECH_1(声音)以及TECH_12(外观设计)。在Ⅱ时期，评论主题主要涉及TECH_8(电池)，其次是TECH_0(处理器/网络/数据传输技术)和TECH_2(游戏)。在Ⅲ时期，评论主题主要涉及TECH_6(摄像功能)以及TECH_8(电池)以及TECH_12(外观设计)，其次是TECH_0(处理器/网络/

数据传输技术）、TECH_1（声音）和 TECH_10（附件）。在Ⅳ时期，评论主题主要涉及 TECH_1（处理器/网络/数据传输技术）以及 TECH_8（电池），其次是 TECH_6（摄像功能） TECH_10（附件）和 TECH_12（外观设计）。在Ⅴ时期，评论主题主要涉及 TECH_2（游戏）、TECH_3（触控/NFC/智能遥控技术）以及 TECH_6（摄像功能），其次是 TECH_0（处理器/网络/数据传输技术）、TECH_1（声音）、TECH_7（存储）、TECH_8（电池）和 TECH_12（外观设计）。

随着时间推移和产品的完善，与摄像功能技术属性相关的评论主题数量相对稳定增长。随着各类手机游戏的推陈出新，用户的休闲娱乐需求随之提高，游戏相关技术属性下的主题数量增长。随着智能遥控、电子配件等产品技术的推广应用，触控/NFC/智能遥控技术属性下的评论主题数量在近期上升。

4.5.2 产品重要技术属性下的主题内容时期演化规律分析研究

以当前重要技术属性 TECH_10（附件）为例，参见公式（2）计算该属性下不同主题内容之间的相似度。由于主题词的提取结果较为凝练，为了更为精确地分析主题内容地时期演化规律，设定主题之间相似度以 7% 作为阈值，相似度大于该阈值则互为相似主题。采用 DyData 可视化平台以桑基图的形式呈现该技术属性下相邻时期相似主题演化关系，分析该技术属性下的主题内容时期演化规律，如图 4 所示：

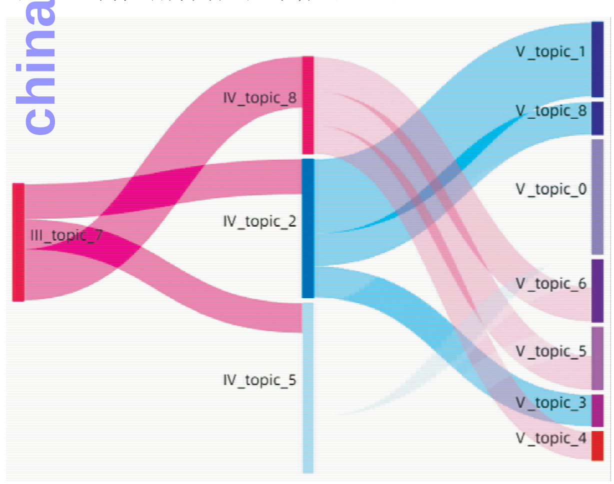
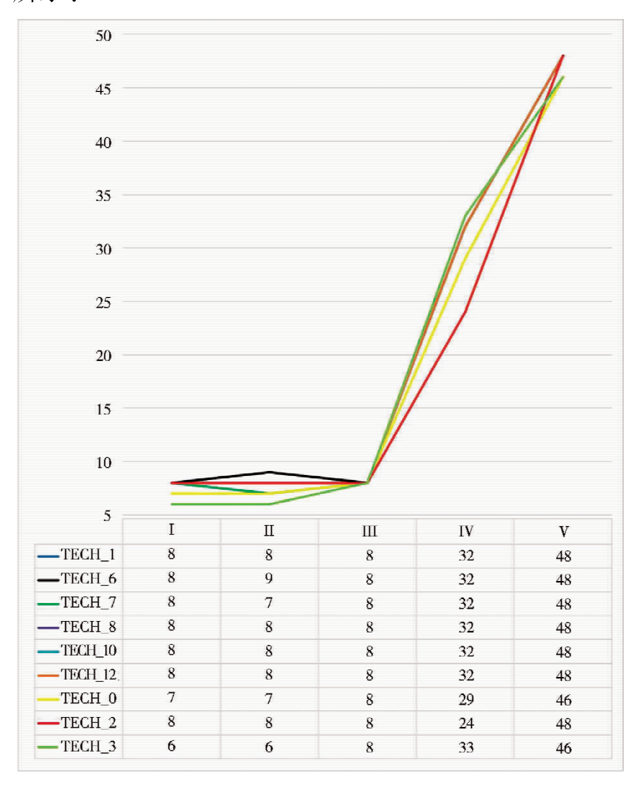


图 4 产品重要技术属性下主题内容时期演化图（以 TECH_10 为例）

从图 4 可以看出，TECH_10（配件）下的主题在研究发展过程中不断交叉、分化与融合。Ⅲ_topic_7 涉及的华为手机配件如手机膜、保护壳以及物流配送技术服务，演化为Ⅳ_topic_2 涉及的小米手机物流配送技术服务、Ⅳ_topic_5 的配件效果及特色以及Ⅳ_topic_8 涉及的苹果手机膜、保护壳以及拍照、充电相关配件的效果与功能。Ⅳ_topic_8 分别与其他主题融合演化为Ⅴ_topic_4 对苹果手机以及华为手机等配件的相关反馈、Ⅴ_topic_5 对苹果手机配件的反馈、Ⅴ_topic_6 对手机充电、拍照等配件的反馈。

4.5.3 产品重要技术属性下的评论数量时期分布规律分析研究

参见公式（2）计算评论与产品技术属性词－技术特征词表之间的相似度。由于预处理过后的评论信息较为凝练，制定的技术属性－技术特征词表较为精确，因此设定评论与属性相似度以 3% 作为阈值，相似度高于阈值则认为是技术属性相关评论，产品重要技术属性下的评论数量时期分布情况如图 5 所示：



	Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ
TECH_1	8	8	8	32	48
TECH_6	8	9	8	32	48
TECH_7	8	7	8	32	48
TECH_8	8	8	8	32	48
TECH_10	8	8	8	32	48
TECH_12	8	8	8	32	48
TECH_0	7	7	8	29	46
TECH_2	8	8	8	24	48
TECH_3	6	6	8	33	46

图 5 产品重要技术属性下评论时期分布情况

从图 5 中可以看出，声音、摄像功能、存储、电池等产品重要技术属性下的评论数量在第Ⅲ到第Ⅴ时期呈现快速增长趋势，这些技术属性下包含了具有未来发展潜力的技术机会。

4.5.4 产品重要技术属性下产品型号的评论时期分布规律分析研究

产品重要技术属性下产品型号的评论时期分布规律，如图 6 所示：

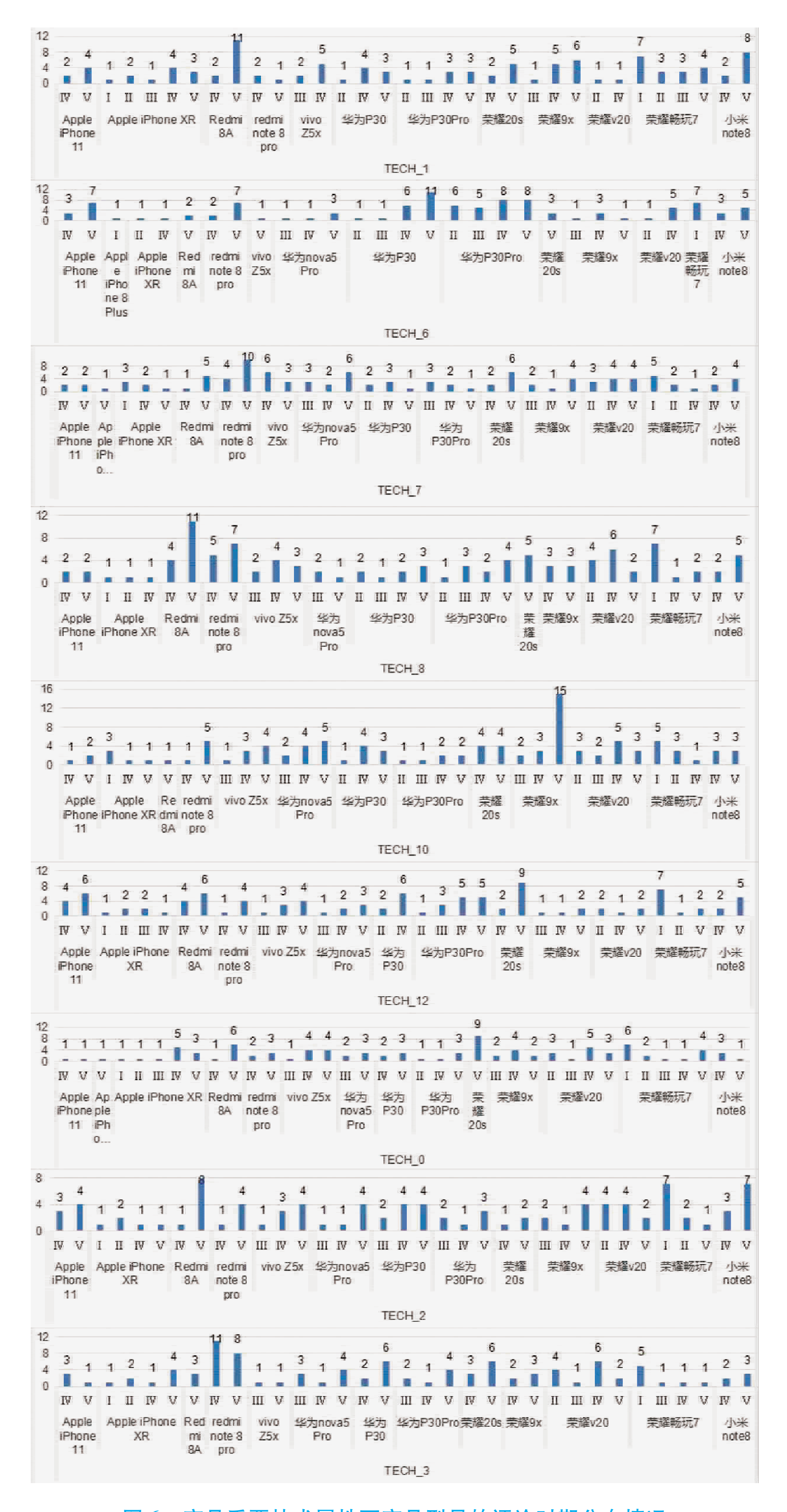


图 6　产品重要技术属性下产品型号的评论时期分布情况

研究者可以根据产品重要技术属性下产品型号的评论时期分布规律，结合网络调研、学术论文、科技报告、专家知识等多元化信息渠道，对特定型号产品的相关属性进行针对性调研分析。例如在第Ⅴ时期，用户对于华为 9X 的 TECH_10(附件)技术属性下评论数量较为突出。经过针对性分析发现，评论反馈主要集中于手机附件的充电插头，插头功率较低，且不支持快充，导致充电时长较长，影响了使用感[21]。研发人员

可以根据技术需求反馈发现技术机会,对产品进行针对性改进,满足用户的技术需求。

5 总结

本研究充分利用产品在线评论数据,提出一种融合评论主题识别与技术属性多维度分析的前瞻技术机会发现研究方法。实验结果表明,首先,从智能手机产品评论中发现的技术机会主要隐含在声音、摄像功能、存储、电池、附件、外观设计等产品重要技术属性中,对不同品牌型号的产品,技术机会有所不同。其次,用户的技术评论主题内容存在相互交叉、渗透与融合的趋势,需要研发人员提高技术研发洞察力,取长补短。最后,尊重并满足评论主题中涉及的对老人等用户群体的产品技术需求是一项重要的技术机会,技术研发人员需要在产品设计与技术研发环节体现更多人文关怀,不断完善产品的科技含量,提高使用感。

一方面,本研究能够从技术需求驱动的视角丰富技术机会发现的研究方法与机理机制,发现隐含在产品评论中的前瞻性技术机会,并提升技术机会发现识别结果准确度,为技术机会发现提供科学客观的研究依据。另一方面,本研究将使技术机会发现研究工作与用户需求端更加紧密结合,为企业精准迎合市场技术需求,把握科技研发机遇动向,优化技术创新资源配置提供及时、前瞻、科学的技术机会研究理论与方法参考。

参考文献:

[1] LI X, XIE Q, JIANG J J. Identifying and monitoring the development trends of emerging Technologies using patent analysis and Twitter data mining: the case of perovskite solar cell Technology [J]. Technological forecasting & social shange, 2019, 146: 687 - 705.
[2] ZHU D H, PORTER A. Automated extraction and visualization of information for Technological intelligence and forecasting [J]. Technological forecasting & social shange, 2002, 69 (5): 495 - 506.
[3] PORTER A, MICHAEL J. DETAMPEL. Technology opportunities analysis[J]. Technological forecasting & social shange, 1995. 49 (3): 237 - 255.
[4] 李保明. 技术机会与技术创新的决策[J]. 科学管理研究, 1990 (5): 61 - 62.
[5] 陈震红, 董俊武. 创业机会的识别过程研究[J]. 科技管理研究, 2005(2): 133 - 136.
[6] 康宇航, 苏敬勤. 技术创新机会的可视化识别——基于专利计量的实证分析[J]. 科学学研究, 2008(4): 695 - 701.
[7] CECERE G, REXHÄUSWE S, SCHULTE P. From less promising to green? technological opportunities and their role in (green) ICT Innovation [J]. Economics of innovation and new technology, 2019, 28(1), 45 - 63.
[8] ORSENIGO L, MALERBA F. Technological regimes and sectoral patterns of innovative activities [J]. Industrial and corporate change, 1997, 6(1): 83 - 117.
[9] 李欣, 谢前前, 洪志生, 等. 基于社会感知分析的新兴技术发展趋势研究——以钙钛矿太阳能电池技术为例[J]. 科技进步与对策, 2018, 35(10): 15 - 24.
[10] 何传启. 新科技革命的预测和解析[J]. 科学通报, 2017, 62 (8): 785 - 798.
[11] 李欣, 黄鲁成. 技术路线图方法探索与实践应用研究——基于文献计量和专利分析视角[J]. 科技进步与对策, 2016, 33(5): 62 - 72.
[12] CHO T S, SHIH H Y. Patent citation network analysis of core and emerging Technologies in Taiwan: 1997 - 2008 [J]. Scientometrics, 2011, 89(3): 795 - 811.
[13] Lee S, Yoon B, Park Y. An approach to discovering new Technology opportunities: keyword-based patent map approach[J]. Technovation, 2008, 29(6): 481 - 497.
[14] 翟东升, 郭程, 张杰, 李登杰. 采用异常检测的技术机会识别方法研究[J]. 现代图书情报技术, 2016(10): 81 - 90.
[15] 王京安, 汤月, 王坤. 基于 Citespace 的技术机会发现研究——以物联网技术发展为例[J]. 现代情报, 2018, 38(2): 130 - 137, 170.
[16] WANG M Y, FANG S C, CHANG Y H. Exploring technological opportunities by mining the gaps between science and TECH-nology: microalgal biofuels[J]. Technological forecasting & social shange, 2015, 92: 182 - 195.
[17] SONG K, KIM K S, LEE S. Discovering new technology opportunities based on patents: text-mining and F-term analysis[J]. Technovation, 2017(60/61): 1 - 14.
[18] 韩晓彤, 刘燕新, 任智军, 等. 基于专利挖掘的技术竞争对手研发方向识别[J]. 科学学与科学技术管理, 2018, 39(2): 23 - 32.
[19] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. The Journal of machine learning research, 2003(3): 993 - 1022.
[20] 张振亚, 王进, 程红梅, 等. 基于余弦相似度的文本空间索引方法研究[J]. 计算机科学, 2005, 032(009): 160 - 163.
[21] 荣耀 9X 充电续航实测, 充电时间让人头疼 但续航表现令人惊喜[EB/OL]. [2020 - 11 - 14]. https://www.sohu.com/a/331268635_120156943.

作者贡献说明:

吴一平:数据收集分析与论文撰写;
白如江:研究命题拟定与论文框架设计;
刘明月:数据分析与论文细节修改;
王效岳:论文框架设计与论文细节修改。

Research on Technology Opportunity Discovery Based on Comment Topic Identification and Multi Dimension Analysis of Technical Attributes

Wu Yiping　Bai Rujiang　Liu Mingyue　Wang Xiaoyue
Institute of Information Management, Shandong University of Technology, Zibo 255049

Abstract: [Purpose/significance] This paper proposed a technology opportunity discovery method which integrated comment topic identification and multi-dimensional analysis of technology attributes, identified technology opportunities from the perspective of technology demand driven, and provided decision-making support for enterprises' forward-looking layout of R & D direction and scientific research management planning. [Method/process] Product online comments were used as the research data source. Firstly, LDA topic model was used to identify the technical topics of comments, and two indicators of technical comment topic strength and topic novelty were proposed to screen out the emerging key technical comment topics. Then, technical attribute words were manually selected from academic papers and technical patents, and high-frequency comment words were obtained through TF-IDF value calculation. Combined with expert knowledge, technical feature words were further selected, and product technical attribute words technical feature words list was constructed. Through the correlation calculation, the technical attributes related to the comments and the topics of the emerging key technology comments were obtained respectively. Finally, this paper proposed an index model to identify important technical attributes of products, and designed a multi-dimensional analysis method to analyze the characteristics of important technical attributes of products, and finally identified the emerging technology opportunities contained in the comment text. [Result/conclusion] The experimental results show that this method can effectively identify technology opportunities prospectively, and provide reference for enterprise product technology R & D management.

Keywords: technology opportunities discovery　technical attributes analysis　subject recognition　comments mining